yang_seonhyeHW16

Seonhye

2/26/2019

```
library(MASS)
library(alr4)
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

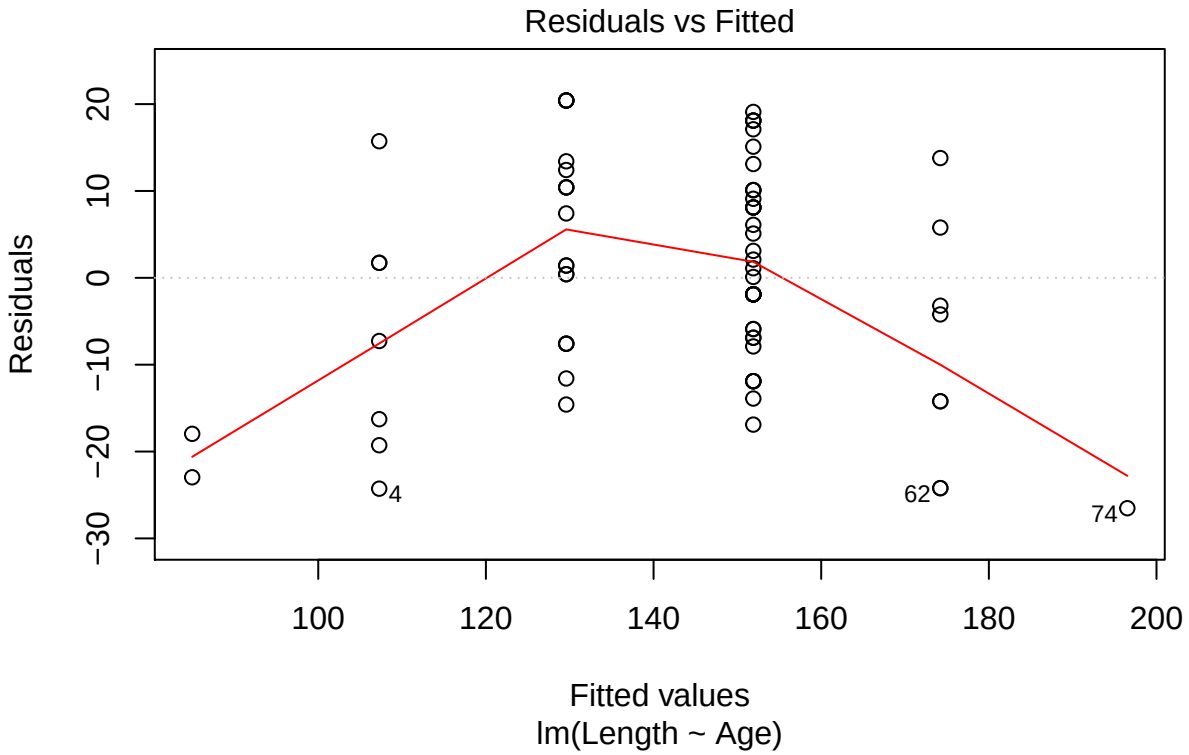
```
## Loading required package: effects
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```
library(data.table)
```

Question 1

```
attach(lakemary, warn.conflicts = F)
lake<- data.frame(lakemary)
fit<- lm(Length~Age, data = lakemary)
plot(fit, which = 1)
abline(fit)
```



THe residual vs fitted shows a quadratic trend. Therefore, a linear model would not be a good fit for the data.

Question 2

We can use this lack of fit test with unknown variance because there are independent replications(the number of observations) at each of the x's. Although it is not possible to expect that the data will have replications,it would be reasonable if replication is built into the study design.

Question 3

```
fitgen = lm(Length ~ factor(Age))
summary(fitgen)$sigma^2
```

```
## [1] 122.3984
```

```
fitlof = anova(fit, fitgen)
fitlof
```

```
## Analysis of Variance Table
##
## Model 1: Length ~ Age
## Model 2: Length ~ factor(Age)
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     76 11892.8
## 2     72  8812.7  4    3080.2 6.2912 0.0002125 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
pure_errors=8812.7/72
pure_errors
```

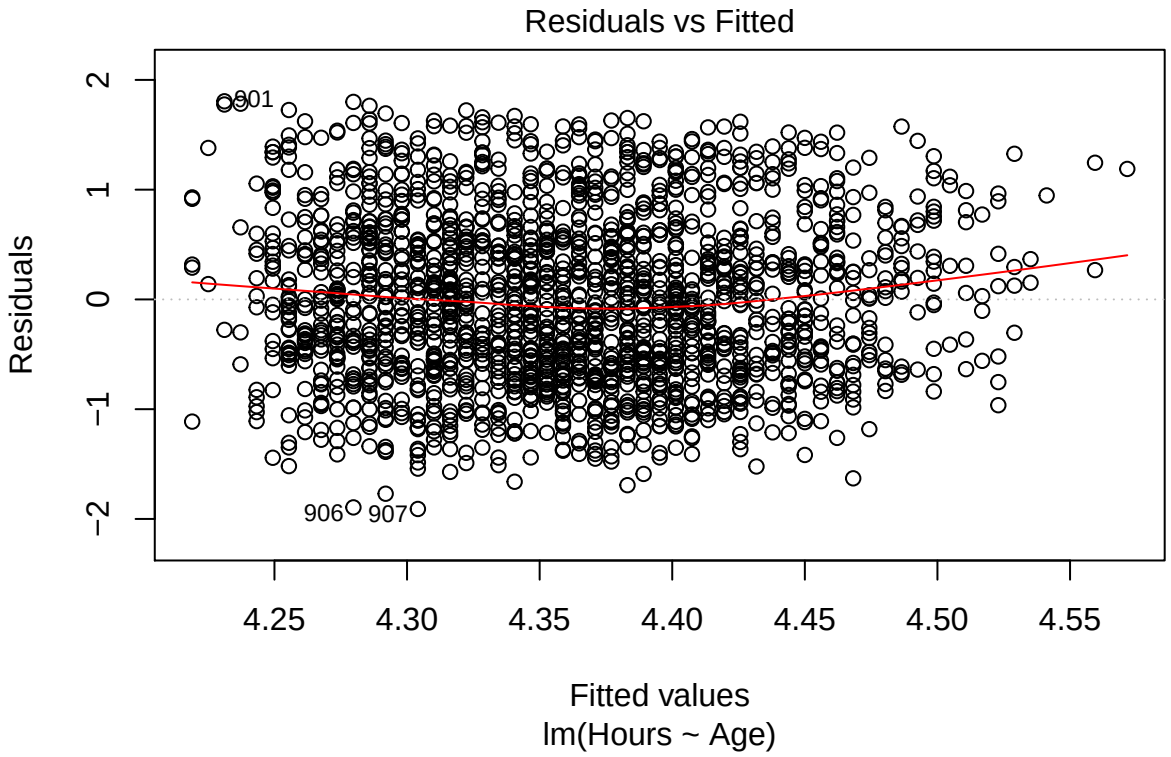
```
## [1] 122.3986
```

Sum of squares for pure error is equal to 8812.7, pure error degrees of freedom is equal to 72, the sum of squares for lack-of-fit is 3080.2 and lack-of-fit degrees of freedom is 4. We find evidence to reject the null hypothesis (linear model) at level $\alpha = 0.05$

Question 4

```
data<- fread("http://users.stat.ufl.edu/~winner/data/napa_marathon_fm2015.csv")
marathon<- data.table("napa_marathon_fm2015.csv" , header = T)
```

```
attach(data, warn.conflicts = F)
fit3<- lm(Hours~Age)
plot(fit3, which = 1)
```




```
fitgen1 = lm(Hours ~ factor(Age))
summary(fitgen1)$sigma^2
```

```
## [1] 0.5636314
```

```
fitlof1 = anova(fit3, fitgen1)
fitlof1
```

```
## Analysis of Variance Table
##
## Model 1: Hours ~ Age
## Model 2: Hours ~ factor(Age)
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1   1880 1072.0
## 2   1826 1029.2 54    42.804 1.4063 0.02848 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can reject the H_0(linear model) at $\alpha = 0.05$

Question 5

```
fit4 = lm(Hours ~ poly(Age, 2, raw = T))
fit4aov = anova(fit4, fitgen1)
fit4aov
```

```
## Analysis of Variance Table
##
## Model 1: Hours ~ poly(Age, 2, raw = T)
## Model 2: Hours ~ factor(Age)
```

```
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1   1879 1059.8
## 2   1826 1029.2 53    30.658 1.0263  0.423
```

Sum of squared of residuals is equal to 1029.2, Sum of squared of lack of fit is equal to 30.658, F-test is 1.0263 and the p-value is 0.423. And we fail to reject the H_0 (linear model) at $\alpha = 0.05$.